How Does This Look? Desirability Methods for Evaluating Visual Design

Edward S. De Guzman and Julie Schiller

Autodesk, Inc., One Market St., Suite 500, San Franciso, California 94105, USA
{Edward.DeGuzman,Julie.Schiller}@autodesk.com

Abstract. Previous studies show that traditional usability evaluation methods can be problematic for collecting feedback on visual design [1]. Desirability studies have been used by usability practitioners to collect feedback on the affective response to interactive systems, but none focus on assessing visual design. We describe the Visual Design Card Sort (VDCS), a desirability research method for collecting visual design feedback from domain experts. We also present a case study where VDCS was used to collect visual design feedback from expert users of Computer-Aided Design (CAD) software. Additionally, we propose a set of analyses for looking at data from desirability studies from different perspectives. Future research is needed to validate some of the assumptions made in designing VDCS and to understand how findings from VDCS compare to other desirability research methods and traditional evaluation techniques.

Keywords: Desirability, visual design, methods, usability testing.

1 Introduction

User-centered design traditionally includes iterative cycles of designing, building, and evaluating. Early iterations of the design may be lower in fidelity (e.g., paper prototypes), but more mature iterations may include a greater degree of interactivity and fidelity of design. In these later iterations, feedback on the *visual design* of the system (e.g., color palette, imagery, typography) becomes increasingly relevant. Traditional usability evaluation methods can be difficult to adapt for accurately assessing visual design in later design iterations. Feedback for visual design is less about the user's ability to accomplish tasks and more about the affective response to a design.

Desirability studies have emerged as an approach for measuring the affective response to design. Desirability research methods collect attitudinal reactions to a design rather than behavioral reactions. The data collected helps designers understand why different design directions alter the emotion of a user experience. Desirability studies can be coupled with traditional usability studies to achieve several benefits:

- In the ISO definition of usability (9241-11) the term is broken down into three components: effectiveness, efficiency, and satisfaction. Considering desirability

C. Stephanidis (Ed.): Posters, Part I, HCII 2011, CCIS 173, pp. 123–127, 2011.

as a component of satisfaction, analyzing and responding to feedback from these studies it can positively impact the overall system's usability.

- First impressions of a system can be positively impacted by designing for usability. Tractinsky [8] observed that a user with a positive first impression of a website based on visual design may be influenced to perceive the website's usability and utility in a similarly positive way. Lindgaard [5] further suggests that negative first impressions, once formed, take more work to change than impressions that start off as positive or even neutral.
- In some domains of interactive systems, there is an observed link between desirability and trust. For example, design quality was found to be among the features that enhance the feeling of trust in e-Commerce websites. [3]

One common approach to conducting a desirability study is to use post-task and post-study questionnaires [4] focused on desirability. Another approach is to use card sorting exercises, a technique traditionally used to inform the structure of a website or product. While several groups have reported case studies of desirability studies [1,2,7,9], we feel there are several research areas worth exploring. First, it is unclear how desirability methods can be adapted to focus on *visual design*. Second, a design space of analysis visualizations of desirability data would help practitioners effectively communicate findings to stakeholders. Finally, desirability research methods need to be optimized to mitigate the costs in time to conduct the research and analyze the data.

In this paper we propose a method, a Visual Design Card Sort (VDCS) for rapid collection of visual design feedback from domain experts. We report on the results of a study using VDCS and present a set of visualizations that provide multiple perspectives on the data. Finally, we propose future directions for analyzing the data and for optimizing the method itself by exploring variables in the study design to iterate the process to be highly efficient and agile.

2 Visual Design Card Sort

In this section we describe the Visual Design Card Sort, a new method for conducting desirability studies to evaluate visual design.

Participants are seated at a table with a monitor and a set of visual design reaction cards (described later in this section). Participants are then shown a set of images, one at a time. Once the image appears on the monitor, the participant has three to five minutes to select between three to five visual design reaction cards that best answers the question, *"How does this look?"* Blank product reaction cards are provided for when the participant wants to volunteer their own word to describe their reaction. After the participant has selected a set of reaction cards, he/she is asked to explain why the card was chosen before moving on to the next image.

This method for assessing desirability differs from the studies mentioned in Related Work in the following ways:

- This method uses Visual Design Reaction Cards, a subset of the Microsoft Product Reaction Cards from the Desirability Toolkit. Two researchers

performed a card filtering exercise to inform and validate the set of cards, only keeping those that are relevant for providing feedback on the visual treatment of a design.

- We assume participants have a minimum degree of expertise in the domain covered by the design. Since participants have only a few minutes to evaluate a design, we suspect that expertise is needed for what Norman [6] calls "behavioral-level processing" to occur (e.g., understanding metaphors and navigation structures, developing appropriate mental models).
- VDCS is done at the start of the research session and feedback is collected after three to five minutes of review. We adapted the Microsoft method [1] to leverage previous work that opinions on aesthetics are made as soon as 50ms [5]. We gave participants significantly more time to review the design so that behavioral-level processing to visual stimuli can occur and influence the reaction guided by visceral processing [6].

3 Case Study

In November 2010, the Visual Design Card Sort was used to collect feedback on a design for a future Web service from Autodesk. Participants were screened against pre-defined criteria: familiarity with one or more of the core Autodesk products and domain expertise from working as an architect, designer, or CAD manager. Six participants were recruited for this study (1 female, 5 male).

Over seven images, the six participants selected 133 Visual Design Reaction Cards. The cards were coded (one researcher coding the results, one researcher validating the coding) into two valences—"positive" and "neutral/negative" based on the rationale for each card choice. 93 cards (70%) were positive and 40 were neutral/negative.

Fig. 1. Word cloud from Visual Design Card Sort case study

4 Discussion

In this section we propose a set of analyses allowing the researcher or stakeholder to look at the data from a desirability study from multiple perspectives.

Word clouds (Figure 1) present the space of words selected while visually showing the importance of certain data points—cards selected more than once appear larger in the cloud. This is effective in presenting a broad overview of the data *for one design* to stakeholders. The Affect Convergence visualization (Figure 2, left) shows the data sorted by valence distribution—the proportion of positive cards selected are plotted above the x-axis (0% to 100%) and the proportion of neutral/negative cards are plotted below (-100% to 0%). This figure shows to what degree participants agreed in having positive or negative reactions to one design and how this convergence of opinion compares across multiple designs. The Affect Consistency visualization (Figure 2, right) highlights responses that are consistent across participants, only showing cards selected *more than once*. This is a visualization adapted from one proposed in [1] to give stakeholders a view of the data that reflects *popular responses* and their corresponding affective valence.

In this section we build upon existing desirability methods research with new directions for visualizing the data collected from desirability studies. We do not advocate a single visualization over the others. Rather, a "scorecard view" showing the set of visualizations for a design provides a nuanced view of the data and helps design teams and stakeholders in comparing and critiquing their designs.

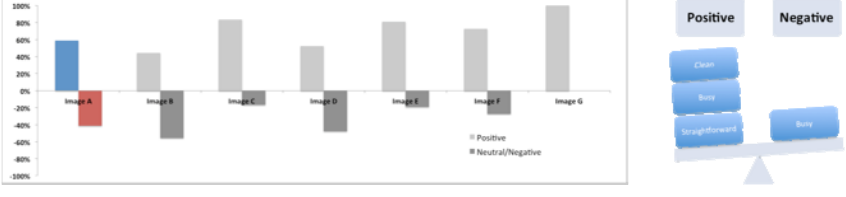

Fig. 2. Affect Convergence (left) and Affect Consistency (right) Visualizations

5 Conclusion and Future Work

Desirability studies are a class of research methods for collecting affective responses to interactive systems. While use of these methods have been documented in case studies and adopted by usability practitioners, there has been little research and innovation in desirability research methods focused exclusively on assessing visual design. Furthermore, a design space of visualizations for presentation of data from desirability studies does not exist. To address some of these open research issues, we present the Visual Design Card Sort (VDCS), a desirability research method focused on collecting feedback from domain experts on visual design. We present the results of using VDCS to assess the visual design of a future Autodesk Web Service, including a set of visualizations for viewing the data from different perspectives.

Future directions include validation of some of the assumptions made which guided the design of this research method. For example, to confirm this method is sensitive to expertise, we propose repeating the case study with *novice* users of CAD programs. We suspect that novice users are more subject to the acquiescence bias and would provide a significantly higher proportion of positive feedback than expert users.

VDCS shows one design for three to five minutes, a relatively longer exposure time than the minimum 50ms reported by Lindgaard for expressing "like" or "dislike" of a website. Further research can investigate the effects of varying the duration of exposure and number of images presented. We suspect that the longer presentation and single stimulus allowed our expert user population to better engage with the visual design and assess its desirability.

Data collected from VDCS can be compared to what is collected from other desirability research methods as well as from traditional evaluation techniques. For example, we propose repeating the case study using the full set of Microsoft Product Reaction Cards and comparing the percentage of positive and negative reaction cards with those found using VDCS. Approximately equal proportions across the two methods would suggest that they are measuring desirability in a similar way.

Finally, the analysis methods described are also opportunities for future research. A principal component analysis (PCA) or other semantic analysis may contribute greater depth to the study than the techniques used in the existing body of work.

References

1. Barnum, C.M., Palmer, L.A.: More than a feeling: understanding the desirability factor in user experience. In: Ext. Abstracts CHI 2010, pp. 4703–4716 (2010)
2. Benedek, J., Miner, T.: Measuring Desirability: New Methods for Evaluating Desirability in a Usability Lab Setting. In: Proc. Usability Professionals Association Conference (2002)
3. Karvonen, K.: The beauty of simplicity. In: Proc. CUU 2000, pp. 85–90 (2000)
4. Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. Int. J. Hum-Comput. Stud. 60(3), 269–298 (2004)
5. Lindgaard, G., Fernandes, G., Dudek, C., Brown, J.: Attention web designers: You have 50 milliseconds to make a good first impression! Behaviour & Information Technology 25, 115–126 (2006)
6. Norman, D.: Emotional Design. Basic Books, New York (2004)
7. Petrie, H., Precious, J.: Measuring user experience of websites: think aloud protocols and an emotion word prompt list. In: Ext. Abstracts CHI 2010, pp. 3673–3678 (2010)
8. Tractinsky, N.: Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In: Proc. CHI 1997, pp. 115–122 (1997)
9. Williams, D., Kelly, G., Anderson, L.: MSN 9: new user-centered desirability methods produce compelling visual design. In: Proc. CHI 2004, pp. 959–974 (2004)